



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Journal Pre-proof

Modelling the persistence of Covid-19 positivity rate in Italy

Antonio Naimoli

PII: S0038-0121(22)00003-9

DOI: <https://doi.org/10.1016/j.seps.2022.101225>

Reference: SEPS 101225

To appear in: *Socio-Economic Planning Sciences*

Received Date: 13 May 2021

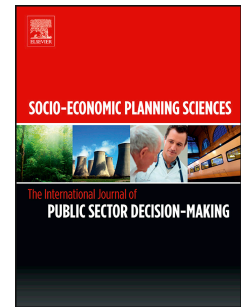
Revised Date: 20 December 2021

Accepted Date: 4 January 2022

Please cite this article as: Naimoli A, Modelling the persistence of Covid-19 positivity rate in Italy, *Socio-Economic Planning Sciences* (2022), doi: <https://doi.org/10.1016/j.seps.2022.101225>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd.



Modelling the persistence of Covid-19 positivity rate in Italy

Antonio Naimoli

Università di Salerno, Dipartimento di Scienze Economiche e Statistiche (DISES)

Abstract

The current Covid-19 pandemic is severely affecting public health and global economies. In this context, accurately predicting its evolution is essential for planning and providing resources effectively. This paper aims at capturing the dynamics of the positivity rate (PPR) of the novel coronavirus using the Heterogeneous Autoregressive (HAR) model. The use of this model is motivated by two main empirical features arising from the analysis of PPR time series: the changing long-run level and the persistent autocorrelation structure. Compared to the most frequently used Autoregressive Integrated Moving Average (ARIMA) models, the HAR is able to reproduce the strong persistence of the data by using components aggregated at different interval sizes, remaining parsimonious and easy to estimate. **The relative merits of the proposed approach are assessed by performing a forecasting study on the Italian dataset. As a robustness check, the analysis of the positivity rate is also conducted by considering the case of the United States.** The ability of the HAR-type models to predict the PPR at different horizons is evaluated through several loss functions, comparing the results with those generated by ARIMA models. The model confidence set is used to test the significance of differences in the predictive performances of the models under analysis. Our findings suggest that HAR-type models significantly outperform ARIMA specifications in terms of forecasting accuracy. **We also find that the PPR could represent an important metric for monitoring the evolution of hospitalizations, as the peak of patients in intensive care units occurs within 12-16 days after the peak in the positivity rate.** This can help governments in planning socio-economic and health policies in advance.

Keywords: HAR, ARIMA, Covid-19, positivity rate, forecasting.

JEL Codes: C58, C22, C53

1. Introduction

Following the China's report (31 December 2019) of a cluster of cases of pneumonia of unknown aetiology (later identified as a new coronavirus Sars-CoV-2) in the city of Wuhan, the World Health Organisation (WHO) declared on 30 January 2020 the novel coronavirus outbreak (Covid-19) a Public Health Emergency of International Concern. On 11 March 2020, the Covid-19 outbreak was identified by the WHO as a global pandemic. **As of 01 December 2020, Covid-19 has infected more than 64.06 million people with 1.54 million global deaths since its emergence. One**

*Corresponding author: Antonio Naimoli, Università di Salerno, Dipartimento di Scienze Economiche e Statistiche (DISES), Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy.

Email address: anaimoli@unisa.it (Antonio Naimoli)

year later, 01 December 2021, the number of confirmed cases has risen to 263.52 million, while the number of confirmed deaths has jumped to 5.23 million¹.

The general understanding of the evolution of the Covid-19 pandemic by researchers and policymakers is often based on data on confirmed cases and deaths. These are the variables that guide public policy in both the introduction and relaxation of non-pharmaceutical interventions, such as masking, social distancing, and other crucial public health measures. However, looking only at the number of cases and deaths can lead to a highly misleading picture of the true scale and spread of the Covid-19 pandemic. These data can only be meaningfully interpreted in conjunction with an accurate understanding of the extent and allocation of virus testing (Hasell et al., 2020). Rapid, reliable, and accurate testing to confirm cases is a prerequisite for successful contact tracing because, without it, infected individuals may remain unidentified and continue to act as sources that support community transmission. Control of the epidemic can be hampered when the percentage of asymptomatic cases is high as in the case of Covid-19 (Oran and Topol, 2020). Thus, when the virus spreads, the only way to identify all infected individuals would be through a universal testing program. Therefore, test positivity is a metric of potential relevance in situations where the percentage of asymptomatic cases is particularly high (Fasina et al., 2021). In addition, according to World Health Organization (2020), before governments can ease restrictions or begin to reopen, the positivity rate (or percent-positive rate) should be 5% or less for at least 14 days. Accordingly, the positivity rate (PPR) is a critical measure as it provides an indication of how prevalent the infection is in the testing area and whether testing levels are keeping up with disease transmission levels.

The Covid-19 pandemic, and the associated stringent preventive measures taken by governments in response to the rapid growth in infections and deaths, have led to unprecedented socio-economic challenges around the world. For example, analyzing the economic connection with Covid-19, Coccia (2021) pointed out that the pandemic and containment policies produced negative effects on economic growth. As a result, the pandemic shocked the global economy, from financial markets where asset prices declined and volatility increased characterizing both the impact and future uncertainty of the pandemic (Baker et al., 2020; Chowdhury et al., 2021; Klose and Tillmann, 2021; Zhang et al., 2020), to the impacts on the supply chain (Gunessee and Subramanian, 2020; Ivanov and Dolgui, 2020; Paul and Chowdhury, 2020a). Italy was the first Western country to experience a Covid-19 emergency with a spiral of infections that placed the country at the top of the international rankings (surpassing China on 19 March 2020 with 41,035 confirmed cases and 3,405 deaths²) and consequently faced large-scale health and socio-economic challenges (Berardi et al., 2020). Several hypotheses have been advanced to explain the rapid spread of the virus within the country, including various geographical, environmental, and socio-economic similarities between Hubei Province in China (where the epidemic broke out) and Northern Italy (Murgante et al., 2020). Several studies have shown that airborne transmission is the dominant route of spread of Covid-19 disease, i.e., the main human-to-human diffusion mechanism (Jayaweera et al., 2020). However, other variables have been identified as potentially contributing to the spread of Covid-19, such as demographic factors (Bertuzzo et al., 2020; Copiello and Grillenzoni, 2020; Diao et al., 2021), environmental and climatic factors, (Bashir et al., 2020; Eslami and Jalili, 2020; Srivastava, 2020) air pollution (Copat et al., 2020; Fattorini and Regoli, 2020; Murgante et al., 2020) and social interactions associated with the mobility of people

¹Johns Hopkins University & Medicine - <https://coronavirus.jhu.edu/>

²Italian Ministry of Health - Daily Bulletin Covid-19 Outbreak in Italy 19/03/2020

and economic activities (Bloise and Tancioni, 2021; Bontempi and Coccia, 2021; Bontempi et al., 2021; Ehlert, 2021). Along these lines, Panarello and Tassinari (2021), making use of a containment index, sanction data, and Google's movement trends across Italian provinces, provided evidence of a deterrent effect on mobility given by the increase in sanction rate and positivity rate among the population.

Overall, the spread of the novel coronavirus has severely affected the integrity of the global economic, social, financial, and behavioral system. Analyzing the relationships between epidemiological and economic models, Verikios (2020) pointed out that because of the greater uncertainty surrounding its nature and the stringent preventive measures taken by governments, Covid-19 is likely to be of longer duration and consequently more severe in its economic effects than previous pandemics. The current outbreak has turned out to be the first extraordinary long-term disruption of the global supply chain (Ivanov, 2020; Ivanov and Dolgui, 2020; Koonin, 2020), and differently from the past, all supply chain players have been severely affected by this pandemic (Chowdhury et al., 2020; Paul and Chowdhury, 2020a; Gunessee and Subramanian, 2020). For example, for some supply chains, demand for necessary items such as personal protective equipment and beauty and personal care products has increased (Ivanov, 2020; Chowdhury et al., 2020; Paul and Chowdhury, 2020a). Conversely, for other industries, such as transportation and manufacturing, demand and supply have decreased dramatically, causing a halt in production (Gray, 2020; Majumdar et al., 2020). In response to the current vulnerability of the entire supply system, several resilience strategies have been suggested to mitigate the impacts of Covid-19 and to recover from the current pandemic (see, e.g. Paul and Chowdhury, 2020a; Taqi et al., 2020; De Silva et al., 2021; Lozano-Diez et al., 2020; Paul and Chowdhury, 2020b; Deaton and Deaton, 2020; Ivanov and Das, 2020; Remko, 2020, among others). The rapid spread of Covid-19 has also significantly affected sustainability, raising several environmental, economic, and social issues (Dente and Hashimoto, 2020; van Barneveld et al., 2020; Sarkis, 2020; Sharma et al., 2020; Queiroz et al., 2020; Rendana et al., 2021; Ibn-Mohammed et al., 2021). However, in this negative scenario, it is worth highlighting some positive aspects mainly related to environmental sustainability: improved air quality, low carbon dioxide and greenhouse gas emissions, decreased energy use and environmental pollution (Dente and Hashimoto, 2020; Ibn-Mohammed et al., 2021; van Barneveld et al., 2020; Sarkis et al., 2020).

In this context, it is necessary to formulate functional planning for the health infrastructure and services in order to curb the spreading of the Covid-19 pandemic. An accurate forecast of the epidemiological trends is essential for health system management and government reform planning. Therefore, to support non-pharmaceutical intervention policies during the Covid-19 outbreak, several models have been proposed for fitting and forecasting the epidemic evolution (see, e.g., Albani et al., 2021; Giordano et al., 2020; Sun et al., 2020).

In the past, the Autoregressive Integrated Moving Average (ARIMA) model has been widely employed for forecasting time series of epidemic diseases. For example, ARIMA models have been successfully applied to estimate the incidence of Severe Acute Respiratory Syndrome (SARS) (Earnest et al., 2005), malaria (Gaudart et al., 2009), tuberculosis (Zheng et al., 2015), influenza viruses (He and Tao, 2018) and brucellosis (Cao et al., 2020). This class of models is also being used to estimate and predict the evolution of the ongoing pandemic. Benvenuto et al. (2020) performed ARIMA model on world data to predict the epidemiological trend of the prevalence and incidence of Covid-2019. In Singh et al. (2020) it was used to predict confirmed cases, deaths, and recoveries for the top 15 countries. Sahai et al. (2020) employed ARIMA models on the daily time series of

total infected cases for US, Brazil, India, Russia and Spain for forecasting the spread of Covid-19. [Monllor et al. \(2020\)](#) applied it to analyze the series of infected persons in China, Italy and Spain, finding a common pattern of disease spread. [Ceylan \(2020\)](#) used ARIMA specifications to predict the epidemiological trend of total confirmed cases of Covid-19 in Italy, Spain, and France.

Most of these papers that rely on ARIMA specifications, model the dynamics of the total number of infected cases, deaths or recoveries. To support policymakers in defining guidelines for the management of health systems, as well as to facilitate the development of plans for economic recovery, this paper investigates the dynamics of the Covid-19 positivity rate, defined as the number of new positive cases divided by the number of total tests. **Our empirical analysis reveals that the positivity rate is characterized by a slowly moving long-run level and a highly persistent autocorrelation structure.**

ARMA models are not well suited to model the long-term behavior of time series. Long memory processes are characterized by a high-order correlation structure indicating persistent dependence between distant observations, implying that the effect of shocks takes a very long time to disappear. The conventional ARMA process is often referred to as a short memory process as it is unable to capture the dynamics of a long memory series. On the other hand, the Autoregressive Fractionally Integrated Moving Average (ARFIMA) process, allowing the order of integration of a series to take on fractional values, provides a useful tool for modelling and forecasting time series with long memory properties ([Baillie, 1996](#)). However, being a fractional integration model, the ARFIMA is not trivial to estimate and lacks a clear economic interpretation. This leads us to introduce a new approach to directly model and forecast the Covid-19 positivity rate. Namely, the time series behavior of the PPR can be adequately captured by the Heterogeneous Autoregressive (HAR) model by [Corsi \(2009\)](#). In this context, the HAR model represents an attractive alternative because of its computational simplicity, ease of interpretation, and remarkable forecasting performance.

Although formally not belonging to the class of long memory models, the HAR model is able to closely mimic the observed long memory behavior by using variables aggregated at different interval sizes. Therefore, differently from the standard ARIMA specifications, HAR-type models are based on an additive cascade of components, from high-frequencies to low-frequencies, allowing to capture both the high degree of persistence (through the long-term component) and short-term dynamics (through the short- and medium-term components) that characterize the PPR behavior. **Along with conventional HAR-type models, we also consider the possibility of selecting relevant lagged components through flexible HAR specifications based on the use of the least absolute shrinkage and selection operator (lasso) ([Tibshirani, 1996](#)) and the adaptive lasso ([Zou, 2006](#)).**

The aim of this paper is to assess the usefulness of the HAR as a new modelling approach to predict the spread of the Covid-19 by capturing the short-, medium- and long-term dynamics of the PPR. Accurate short- and long-term predictions of the PPR can be essential both for developing non-pharmaceutical strategic planning by policymakers to address the current health emergency and for shaping new policies to overcome the severe negative impacts experienced by businesses and supply chains because of the pandemic. The positivity rate measures both the severity of the outbreak and the limitations of testing. That is, the PPR is a useful measure of whether sufficient testing has been done and what the current level of SARS-CoV-2 transmission is in the community. Therefore, this approach could provide a useful tool for both monitoring the spread of the virus and guiding policymakers to undertake actions to curb the spread of the

disease. HAR-type models prove to be particularly useful as, on the one hand, they are able to adequately predict the short-, medium- and long-term trend of the positivity rate, and, on the other hand, the lasso-based HAR specifications are completely data-driven, thus reducing uncertainty in the choice of predictor lags. Therefore, the proposed approach provides a reliable tool that simplifies the decision-making process by moving towards a single data-driven direction.

The profitability of the HAR approach in forecasting the Covid-19 positivity rate is evaluated through an application on the Italian dataset, as it was the first European country to be seriously affected by the pandemic. On 30 March 2020, more than 101 thousand people were positive to Covid-19³. The empirical application shows that HAR-type models outperform the most popular ARIMA models revealing that the improvements are especially significant for longer forecast horizons as detected by the model confidence set (MCS) (Hansen et al., 2011). These findings are also confirmed by analyzing the positivity rate of the United States, which was considered as a reference country to pursue a robustness check of the proposed approach. **Finally, the positivity rate exhibits predictive ability with respect to hospitalizations.**

The remainder of the paper is organized as follows. Section 2 presents the Heterogeneous Autoregressive model and its lasso-based extensions. Section 3 describes the data and the main non-pharmaceutical measures adopted by the Italian government. Section 4 illustrates the results of the empirical study and robustness checks. Section 5 presents a broader discussion of the positivity rate along with some limitations and caveats. Finally, Section 6 summarizes the findings with concluding remarks.

2. Heterogeneous autoregressive models

Inspired by the Heterogeneous Market Hypothesis (Müller et al., 1993) and the asymmetric propagation of volatility between long-term and short-term horizons, Corsi (2009) proposed the HAR model to parsimoniously capture the strong persistence typically observed in Realized Volatility (RV) (Andersen and Bollerslev, 1998) by the sum of lagged RV components aggregated over different interval sizes.

The HAR model is commonly used in modelling the dynamics of financial volatility as it is able to reproduce the main stylized facts of financial data such as the long memory and asymmetric propagation of volatility over time. In most empirical applications, the HAR model is specified as an additive cascade of three volatility components aggregated over different time intervals, that is daily, weekly and monthly, which implies a fixed (1,5,22) lag structure. However, the structure (1,5,22) may not fully reflect the characteristics of the data. Thus, determining the optimal lag structure of the HAR could significantly improve the predictive ability of the model. In this direction, Audrino and Knaus (2016) showed that the HAR-implied lag structure can be recovered asymptotically by the lasso only if the HAR is the underlying data generating process (DGP). On the other hand, differently from Audrino and Knaus (2016) who employed the lasso on the Autoregressive (AR) framework, Audrino et al. (2019) referred to the adaptive lasso to investigate whether the lag structure implied by the HAR can be identified. However, their results highlight the difficulty of outperforming the forecast performance of the standard HAR model based on the daily, weekly and monthly components.

The HAR model can be easily estimated by ordinary least squares (OLS), showing remarkable good volatility forecasting performance. Therefore, it is widely used to model the dynamics of

³Italian Ministry of Health - Daily Bulletin Covid-19 Outbreak in Italy 30/03/2020

RV, but it has never been used to predict the evolution of pandemics. In light of its inherent characteristics, to capture the slowly decaying autocorrelation structure (also known as *long memory*) of the Covid-19 positivity rate series, we propose to apply the HAR modelling approach.

Let PPR_t be the positivity rate at time t . The HAR model for the h -step-ahead daily PPR_t can be specified as

$$PPR_{t+h} = \beta_0 + \beta_1 PPR_t^{(1)} + \beta_2 PPR_t^{(7)} + \beta_3 PPR_t^{(28)} + \varepsilon_{t+h}, \quad (1)$$

where $PPR_t^{(k)} = k^{-1} \sum_{j=1}^k PPR_{t-j}$ is the k -period average of daily PPR and ε_t is a zero mean innovation.

This specification, substantially, states that tomorrow's PPR is a weighted sum of daily, weekly and monthly averages of PPRs that can be characterized by different dynamics of virus infection and transmission over time. For example, to take into account the different periods of time between infection and development of clinical symptoms as well as the transmission periods, the model in Equation 1 can be further extended as

$$PPR_{t+h} = \beta_0 + \beta_1 PPR_t^{(1)} + \beta_2 PPR_t^{(7)} + \beta_3 PPR_t^{(14)} + \beta_4 PPR_t^{(21)} + \beta_5 PPR_t^{(28)} + \varepsilon_{t+h}. \quad (2)$$

Thus, the HAR model is parsimonious, it allows to approximate long memory in a very simple way and it can be consistently estimated by OLS.

In this context, it becomes crucial to define the lag structure and the maximum order of the model. It is worth noting that the HAR can be represented as a constrained AR(p) model (Corsi, 2009). Considering the HAR process introduced in Equation 1, we can write it as a restricted AR(28) process, namely

$$PPR_{t+1} = \phi_0 + \sum_{i=1}^{28} \phi_i PPR_{t-i+1}^{(1)} + \epsilon_{t+1} \quad (3)$$

where

$$\phi_i = \begin{cases} \beta_1 + \frac{1}{7}\beta_2 + \frac{1}{28}\beta_3 & \text{for } i = 1 \\ \frac{1}{7}\beta_2 + \frac{1}{28}\beta_3 & \text{for } i = 2, \dots, 7 \\ \frac{1}{28}\beta_3 & \text{for } i = 8, \dots, 28 \end{cases} \quad (4)$$

In contrast to the fixed daily-weekly-monthly time scale, extensions of the HAR model have been proposed to allow for potentially different predictive information arising from a different lag structure. In this direction, to investigate whether a more general lag structure provides more accurate predictions than the fixed (1,5,22) lag index, Audrino and Knaus (2016) and Audrino et al. (2019) compared the standard HAR model to a lasso-based method.

Let the daily PPR_t be denoted by x_t , with $(x_t, \dots, x_{t-p+1})'$ the predictor variables. Then, the lasso⁴ (Tibshirani, 1996) estimator of the AR(p) model

$$x_{t+1} = c + \sum_{i=1}^p \phi_i x_{t-i+1} + \varepsilon_{t+1}$$

⁴The lasso (least absolute shrinkage and selection operator) method aims to select a subset of important covariates by shrinking the coefficients of redundant ones towards zero.

is obtained as

$$\phi^{lasso} = \underset{\phi}{\operatorname{argmin}} \left\{ \sum_{t=p}^T \left(x_{t+1} - \sum_{i=1}^p \phi_i x_{t-i+1} \right)^2 + \lambda \sum_{i=1}^p |\phi_i| \right\} \quad (5)$$

where $\lambda \geq 0$ is the tuning parameter which controls the strictness of the penalty term, with $\lambda = 0$ leading to the OLS estimator. The solution for the constant c is $\hat{c} = \bar{x}$, that is zero for demeaned data.

It has been shown that the lasso suffers from some drawbacks due to the lack of oracle properties. On the other hand, the adaptive lasso (Zou, 2006) estimator fulfils the oracle property in the sense introduced by Fan and Li (2001), as it allows asymptotically consistent and efficient variables selection and provides asymptotically unbiased and normally distributed estimates of the non-zero coefficients.

The adaptive lasso estimator is given by

$$\phi^{alasso} = \underset{\phi}{\operatorname{argmin}} \left\{ \sum_{t=p}^T \left(x_{t+1} - \sum_{i=1}^p \phi_i x_{t-i+1} \right)^2 + \lambda \sum_{i=1}^p \lambda_i |\phi_i| \right\} \quad (6)$$

where the weights λ_i can be computed as the inverse of the absolute value of the corresponding preliminary ridge regression or OLS estimator. The ordinary lasso is obtained as a special case for $\lambda_i = 1, \forall i = 1, \dots, p$.

The K -fold cross-validation is used to determine the optimal tuning parameter λ . Specifically, the data are randomly divided into K groups (G_1, \dots, G_K) and for each group the mean squared error is estimated on the validation set by

$$e_k(\lambda) = \sum_{t \in G_k} (x_t - \hat{x}_t)^2. \quad (7)$$

For each tuning parameter value, the average error over all folds is computed as

$$CV(\lambda) = \frac{1}{T} \sum_{k=1}^K e_k(\lambda) = \frac{1}{T} \sum_{k=1}^K \sum_{t \in G_k} (x_t - \hat{x}_t)^2 \quad (8)$$

and thus the optimal λ is chosen by minimizing the $CV(\lambda)$ function, that is

$$\hat{\lambda}_{CV} = \underset{\lambda}{\operatorname{argmin}} CV(\lambda). \quad (9)$$

The Flexible HAR (FHAR) of Audrino et al. (2019) can be estimated by applying the adaptive lasso procedure to select the active terms to be included in the model considering the following equation

$$PPR_{t+1} = \beta_0 + \sum_{i=1}^p \beta_i \left(\sum_{j=1}^i \frac{1}{i} PPR_{t-j+1}^{(1)} \right) + \varepsilon_{t+1}. \quad (10)$$

Motivated by these theoretical developments, this paper aims to apply the HAR and its lasso extensions to Covid-19 data to predict its spread and provide a good alternative to other models that have been proposed to study the dynamics of the ongoing pandemic.

3. Data description and management of the 2020 Covid-19 pandemic in Italy

The data used in this paper refer to the daily number of confirmed Covid-19 cases and daily total tests in Italy, between 24 February and 20 December 2020 for a total of 301 days. The data have been downloaded from the official Civil Protection Department website⁵ - Presidency of the Council of Ministers.

Table 1: Summary statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.Dev.
Daily new positive cases	78	384	1,501	6,498	5,560	40,902	10,143.22
Daily tests	964	41,867	61,725	84,259	108,019	254,908	62,825.32
Daily positivity rate	0.23	0.90	2.39	6.69	11.61	46.21	7.61

Table 1 provides summary statistics for the new positive cases, number of tests (swabs) performed and PPR at a daily level in Italy for the full sample period. The occurrence of new cases of a disease developing in a population over a period of time, also known as “incidence” in epidemiology, can be used to map the frequency with which Covid-19 develops in a community⁶. These peaked at 40,902 during the *second wave* (October - November 2020) along with the number of tests performed in a single day, 254,908. On the other hand, the positivity rate touched 46.21% during the *first wave* (February - March 2020), reaching its minimum of 0.23% in June during the *Phase 2*, characterized by an easing of previously adopted restrictive measures. A possible explanation for the high positivity rate in the very early phase of the epidemic is that on 25 February 2020, the Italian Ministry of Health issued more stringent testing policies. That is, testing was prioritized for patients with more severe clinical symptoms who were suspected of having Covid-19 and required hospitalization. Consequently, testing was limited for people who were asymptomatic or had mild symptoms. This strategy inevitably led to a high percentage of positive tests (Onder et al., 2020).

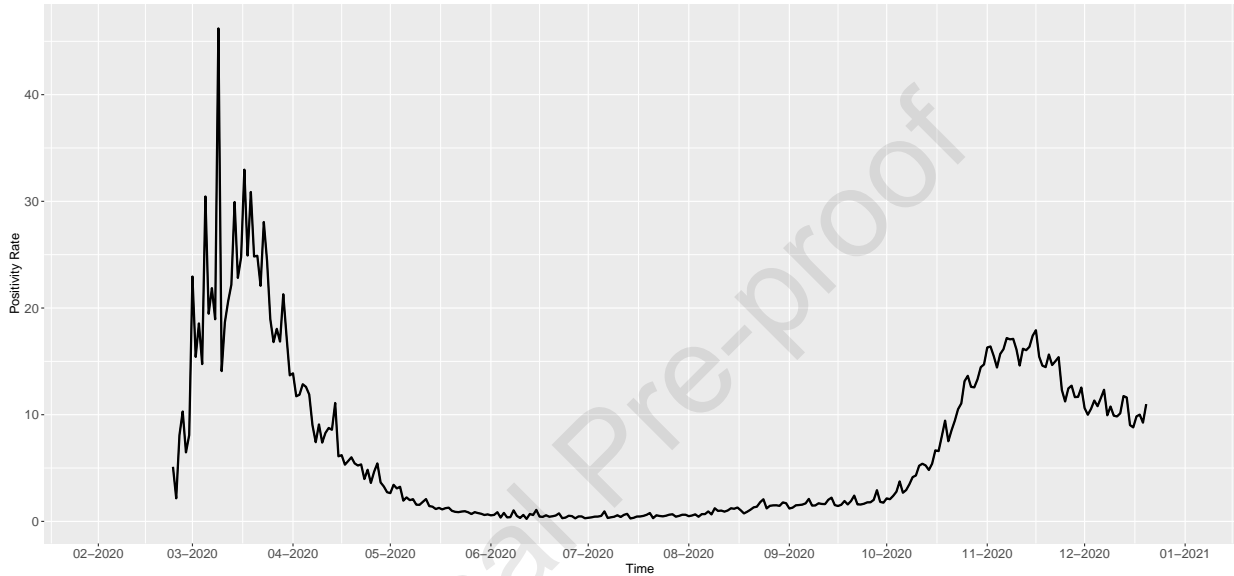
Figure 1 displays the time series of the daily positivity rate given by $(\text{new positive cases} / \text{total tests}) \times 100$ for Italy. It also shows that the PPR exhibits persistence, i.e. large changes in the positivity rate are often followed by other large changes and small changes are often followed by small changes. The presence of long memory can be identified by a data-driven empirical approach in terms of the persistence of observed autocorrelations. This feature is highlighted in Figure 2 which displays the daily and weekly autocorrelation functions for the PPR up to lag 40. The correlograms show that the autocorrelations exhibit a clear pattern of slow decay and persistence. In particular, the sample autocorrelations reveal the presence of a hyperbolic decay rate, which is much slower than the usual geometric rate associated with stationary ARMA processes. Also, it was not possible to reject the null hypothesis of a unit root at the 5% level using

⁵<http://www.protezionecivile.gov.it/home>

⁶Incidence differs from prevalence because incidence is limited only to new cases during a specified time period, whereas prevalence includes all cases (both new and existing) in the population either at a specific point in time (the point prevalence) or over a specified time period (the period prevalence). The incidence can be measured as a proportion or as a rate. For the incidence proportion, the numerator is the number of new cases of a disease or condition that occur during a given time period, while the denominator is the total population at risk during the defined study period. For the incidence rate, the numerator is still the number of new cases observed over a given period of time, but the denominator is now the sum of the person-time at risk (for further details see, e.g., Porta, 2014).

the Augmented Dickey-Fuller test. However, the classical trend stationary $I(0)$ and unit roots $I(1)$ representations may be too restrictive with respect to the low-frequency dynamics of the series. This compelling evidence of long memory, i.e. the historical PPRs have a persistent impact on the future PPR, suggests that the Covid-19 positivity rate can be adequately modelled through HAR specifications.

Figure 1: Positivity rate for Italy



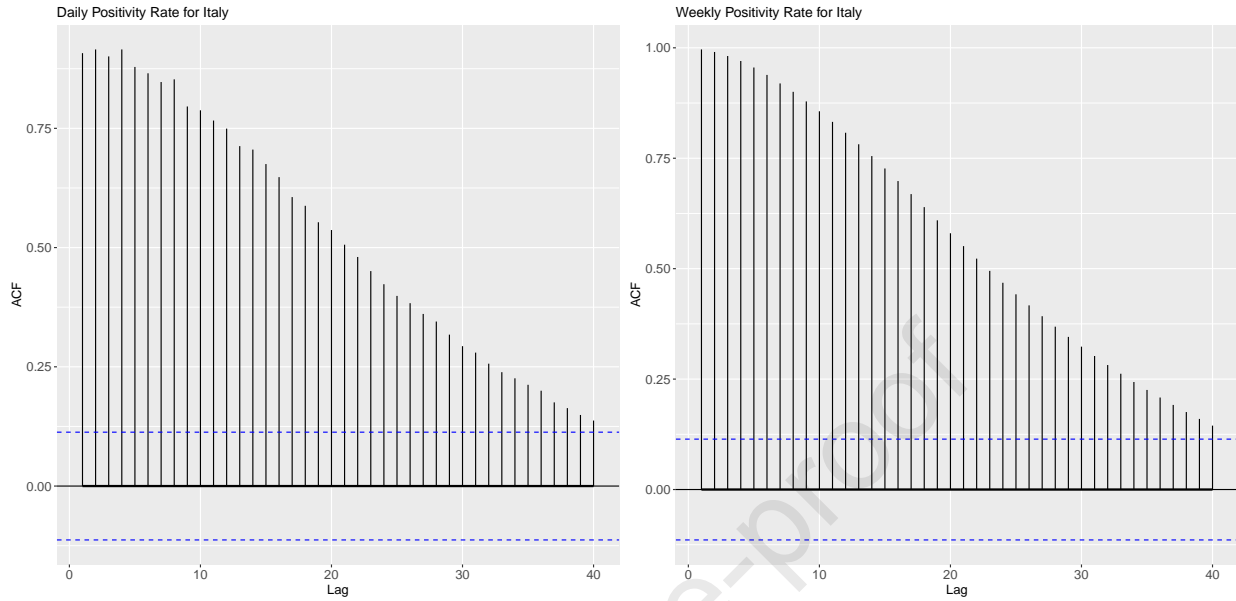
Time series of daily positivity rate for Italy. Sample period: 24 February 2020 – 20 December 2020.

To better understand the dynamics of the contagion and the importance of PPR behavior in guiding decisions about reopening schools and businesses, we briefly report the main measures taken by the Italian government to contain the epidemic⁷. For a more extensive analysis of policy interventions implemented by the Italian government and their impact on health and non-health outcomes, see Berardi et al. (2020).

The Italian government confirmed the first cases of the disease in the country on 30 January 2020, when the novel coronavirus was detected in two Chinese tourists while visiting Italy. On request of the Italian Health Authorities, all flights to/from People's Republic of China (PRC) including Hong Kong, Macao and Tapei were suspended. Once the first internal outbreak was discovered, one of the first measures adopted was the quarantine of 11 municipalities in Northern Italy located in Lombardy and Veneto.

On 23 February, the Council of Ministers decreed the total closure of the municipalities with active outbreaks. This is also confirmed by Figure 1, that shows how the positivity rate continues to grow after 23 February, reaching its maximum of 46.21% on 09 March. During this time, there was a succession of different measures aimed at containing the epidemic, and on 09 March, the so-called *Phase 1*, began with the country being locked down until 03 May 2020. Italy was the first country to implement a national quarantine due to the 2020 coronavirus outbreak. As a result,

⁷<http://www.governo.it/it/coronavirus-misure-del-governo>

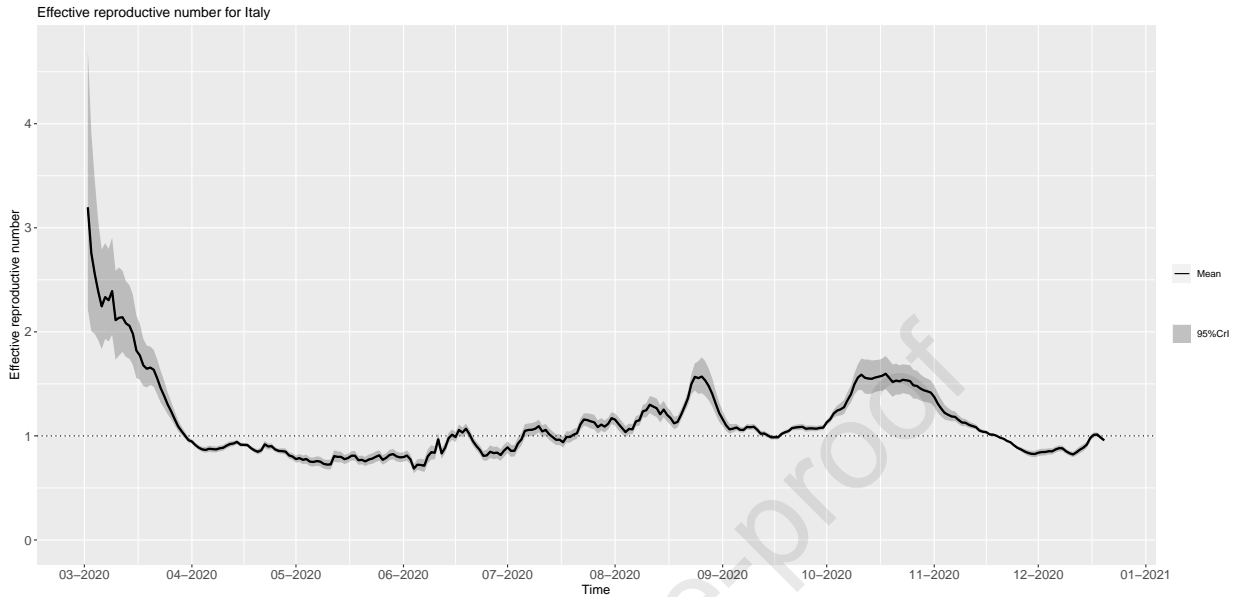
Figure 2: Autocorrelation function for the daily (left) and weekly (right) positivity rate for Italy

the positive percentage rate began to slowly decline towards zero and on 26 April, the then-Prime Minister Giuseppe Conte announced the so-called *Phase 2*, that would start from 04 May. *Phase 2* was characterized by a gradual relaxation of previous containment measures. Italy therefore tried to restart by reopening bars, restaurants and shops. All while observing the new safety rules, ranging from social distancing to the use of face masks. As it can clearly be seen from [Figure 1](#), the infection curve tends to flatten out and so from 15 June (end of *Phase 2*) to 07 October, *Phase 3* of coexistence with Covid-19 began.

Following the rise of the epidemic curve in the autumn, renewed restrictions were progressively introduced, mainly concerning commercial and private activities rather than restricting movement. This led to the *second wave* of the pandemic, with the positivity rate rising and new restrictive measures being introduced between 08 October and 05 November. Starting from 03 November 2020, the Regions and Autonomous Provinces of Trento and Bolzano have been classified into three areas, namely red, orange and yellow, according to the degree of risk, for which specific restrictive measures were envisaged. This classification is based on the ordinances issued by the Ministry of Health.

To mitigate the effects of Covid-19 and make appropriate health, economic and social system decisions, it is crucial to understand the pandemic evolution. The effective reproductive number (R_t) is a parameter that has been widely used to measure the transmissibility of the ongoing epidemic infection. Therefore, as a further tool to understand the actions enacted by the Italian government to counter the spread of the coronavirus, the R_t is also estimated. The R_t is a fundamental epidemiological parameter that characterizes the temporal dynamics of infectious disease as measuring the average number of secondary cases caused by an infected individual in a population composed of both susceptible and non-susceptible individuals ([Wallinga and Teunis, 2004](#); [Cori et al., 2013](#); [Delamater et al., 2019](#)).

The Covid-19 R_t has been estimated by the [Cori et al. \(2013\)](#) approach, using the *EpiEstim*

Figure 3: Effective reproductive number for Italy

Estimates of the weekly effective reproductive number (R_t , solid line) and 95% credible interval (grey area) during the coronavirus disease 2019 outbreak in Italy. Sample period: 24 February 2020 – 20 December 2020.

package (Cori, 2020) and R Core Team (2020) software. To explicitly take into account the uncertainty in the serial interval (SI) distribution, the mean and standard deviation of the SI (time interval between the onset of symptoms in the primary and secondary cases) are allowed to vary according to truncated normal distributions, employing parameters estimated from existing studies (see, Liu et al. (2020); Nishiura et al. (2020) and Du et al. (2020), among others). Therefore, the R_t was estimated on sliding weekly windows, with values drawn from a Gamma distribution, with the mean and variance sampled from 1,000 truncated normal distributions for which we used an average mean serial interval of 4.8 days ($sd = 2.3$, $min = 3.8$, $max = 5.8$), and an average standard deviation of 2.3 days ($sd = 2.3$, $min = 1.3$, $max = 3.3$).

The resulting weekly R_t series is reported in Figure 3. It clearly shows the peak of transmissibility during the *first wave* (February-March 2020), while between April and around mid-June the R_t index remains below one, indicating that the spread of infection is decreasing. However, with the start of *Phase 3* (15 June), a slight but steady increase in the national transmission index was noted, reaching a summer peak in the period 15 August - 31 August where R_t reached 1.59 (95% CI: 1.41 - 1.75) with an incidence calculated as daily cases. It is worth noting that the estimates have a large stochastic variability, especially with regard to the summer period, which is overall characterized by a small number of cases.

Because of the increase in the number of infections, on 16 August 2020, the Minister of Health Roberto Speranza signed an ordinance ordering the closure of discos and dance halls and making it compulsory, from 6 p.m. to 6 a.m., to wear masks even in public spaces. In September, with the start of the new school year, classroom activities resumed. To reduce the risk of infection, school staff were allowed to undergo serological testing. Finally, during the *second wave* (October-November 2020) an R_t index above 1.5 is recorded for most of October, falling below 1 from mid-

November until the end of December. As of 20 December 2020, a total of 622,760 people tested positive for Covid-19, with 68,799 deaths and 1,261,626 patients discharged/healed, nationwide.

4. Empirical Analysis

In this section, we conduct several empirical studies to compare the in-sample and out-of-sample performance of the HAR with the commonly used ARMA models. The data employed for the empirical analysis consist of the daily Covid-19 positivity rate recorded in Italy between 24 February and 20 December 2020, with a full-sample period of 301 days.

Regarding the Flexible-HAR (FHAR) based on the adaptive lasso, following [Audrino et al. \(2019\)](#), the maximum lag order is set to $p = 50$, while the tuning parameter λ is chosen by five-fold cross-validation⁸. The weights λ_i in the adaptive lasso are calibrated as the inverse of the absolute value of the corresponding preliminary ridge regression estimator ([Audrino et al., 2019](#)).

We also estimate a Flexible-HAR based on the lasso method to select appropriate HAR lag length, resulting in the Lasso-HAR (LHAR) model. All the lasso estimates are obtained using the *glmnet* package ([Friedman et al., 2010](#)).

For the purpose of comparing the in-sample and out-of-sample performance of the analyzed models, we consider the following loss functions ([Patton, 2011](#)):

$$\begin{aligned}
 MSE &= T^{-1} \sum_{t=1}^T (x_t - \hat{x}_t)^2 & MAE &= T^{-1} \sum_{t=1}^T |x_t - \hat{x}_t| \\
 MSE_{log} &= T^{-1} \sum_{t=1}^T (\log(x_t) - \log(\hat{x}_t))^2 & MAE_{log} &= T^{-1} \sum_{t=1}^T |\log(x_t) - \log(\hat{x}_t)| \\
 MSE_{sd} &= T^{-1} \sum_{t=1}^T \left(\sqrt{x_t} - \sqrt{\hat{x}_t} \right)^2 & MAE_{sd} &= T^{-1} \sum_{t=1}^T \left| \sqrt{x_t} - \sqrt{\hat{x}_t} \right| \\
 MSE_{prop} &= T^{-1} \sum_{t=1}^T \left(\frac{x_t}{\hat{x}_t} - 1 \right)^2 & MAE_{prop} &= T^{-1} \sum_{t=1}^T \left| \frac{x_t}{\hat{x}_t} - 1 \right| \\
 QLIKE &= T^{-1} \sum_{t=1}^T \left(\log(\hat{x}_t) + \frac{x_t}{\hat{x}_t} \right)
 \end{aligned}$$

where x_t is the PPR_t and \hat{x}_t is the prediction obtained by the HAR-type or ARIMA-type models.

In addition, we further assess the significance of differences in forecasting performance of all competing models by means of the MCS ([Hansen et al., 2011](#)). The MCS relies on a sequence of statistical tests to identify, at a certain confidence level $(1 - \alpha)$, the set of superior models with respect to some appropriately-chosen measures of predictive ability. The MCS p -values are obtained by 5,000 bootstrap resamples generated by a block-bootstrap procedure, estimating the optimal block length through the method described in [Patton et al. \(2009\)](#).

⁸Our findings are robust to alternative choices of p .

4.1. In-sample results

Table 2 shows the results of the model comparison in terms of in-sample accuracy. In particular, we compare 15 ARIMA specifications with the HAR based on different lag structures, together with LHAR and FHAR. To determine whether differencing in the PPR series is required, we use the Augmented Dickey-Fuller (ADF) test, suggesting that one difference is needed to make the data stationary. This leads to using ARIMA($p,1,q$) specifications⁹.

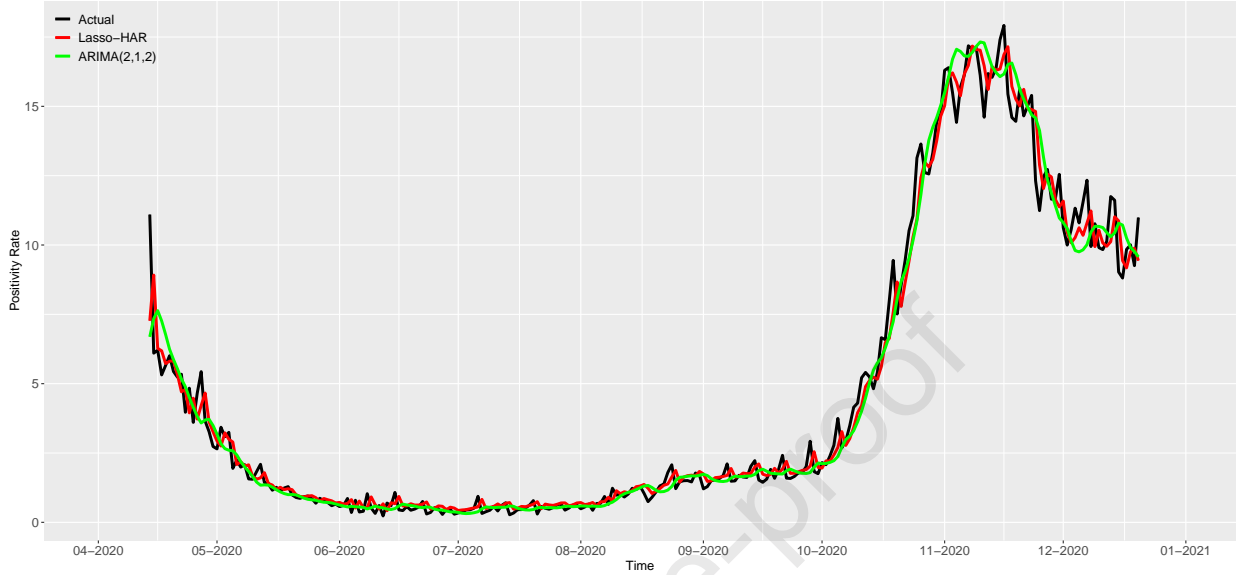
Table 2: In-sample model comparison.

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	0.5317	0.1336	0.1953	0.1931	0.6574	0.0282	0.0691	0.0694	1.8096
ARIMA(2,1,0)	0.5357	0.1324	0.1882	0.1863	0.6979	0.0287	0.0631	0.0627	1.8068
ARIMA(3,1,0)	0.5505	0.1350	0.1889	0.1889	0.7837	0.0315	0.0647	0.0679	1.8081
ARIMA(0,1,1)	0.5263	0.1296	0.1827	0.1807	0.6878	0.0283	0.0609	0.0608	1.8058
ARIMA(0,1,2)	0.5589	0.1367	0.1915	0.1912	0.7891	0.0314	0.0640	0.0664	1.8077
ARIMA(0,1,3)	0.5716	0.1393	0.1946	0.1976	0.8328	0.0328	0.0654	0.0740	1.8093
ARIMA(1,1,1)	0.5413	0.1330	0.1872	0.1856	0.7296	0.0295	0.0620	0.0626	1.8065
ARIMA(1,1,2)	0.5252	0.1301	0.1851	0.1883	0.6919	0.0283	0.0603	0.0685	1.8068
ARIMA(1,1,3)	0.5284	0.1307	0.1857	0.1891	0.7011	0.0286	0.0607	0.0694	1.8070
ARIMA(2,1,1)	0.5428	0.1331	0.1869	0.1857	0.7400	0.0298	0.0617	0.0629	1.8064
ARIMA(2,1,2)	0.4708	0.1169	0.1691	0.1747	0.5997	0.0244	0.0556	0.0672	1.8049
ARIMA(2,1,3)	0.4698	0.1185	0.1745	0.1810	0.5776	0.0242	0.0581	0.0727	1.8064
ARIMA(3,1,1)	0.5478	0.1347	0.1889	0.1889	0.7727	0.0312	0.0649	0.0684	1.8082
ARIMA(3,1,2)	0.5437	0.1366	0.1978	0.1982	0.7413	0.0312	0.0736	0.0788	1.8123
ARIMA(3,1,3)	0.4673	0.1177	0.1729	0.1790	0.5736	0.0240	0.0575	0.0715	1.8060
HAR(1,7,14)	0.4593	0.1222	0.1872	0.1674	0.4711	0.0240	0.0743	0.0504	1.8081
HAR(1,7,21)	0.4674	0.1300	0.2076	0.1798	0.4796	0.0264	0.0900	0.0561	1.8139
HAR(1,7,28)	0.4835	0.1320	0.2073	0.1799	0.5079	0.0274	0.0909	0.0562	1.8141
HAR(1,7,14,21)	0.5008	0.1413	0.2276	0.1933	0.5285	0.0303	0.1060	0.0628	1.8199
HAR(1,7,14,28)	0.4539	0.1209	0.1852	0.1665	0.4567	0.0232	0.0718	0.0498	1.8072
HAR(1,7,14,21,28)	0.4432	0.1157	0.1731	0.1589	0.4572	0.0223	0.0641	0.0485	1.8046
LHAR(1,7,19,25)	0.4255	0.1118	0.1693	0.1563	0.4460	0.0218	0.0645	0.0490	1.8048
FHAR(1,7,27)	0.4410	0.1149	0.1726	0.1590	0.4796	0.0231	0.0675	0.0509	1.8059

The table reports the average values of the different loss functions for the models under analysis. The lowest value of the loss in each column is displayed in bold. The sample runs from 24 February 2020 to 20 December 2020. Note that for LHAR and FHAR the lags are not imposed, but the selected lag structure allowed by the lasso and adaptive lasso methods is reported, respectively.

The empirical results in Table 2 highlight that the selected lag structure by the adaptive lasso for the FHAR is (1,7,27), which is in line with the canonical daily-weekly-monthly lag structure of the standard HAR model, while the lasso suggests using an additional biweekly/triweekly

⁹The values of p and q chosen by minimizing the Corrected Akaike's Information Criterion (AIC_c) are $p = 3$ and $q = 3$. However, in our analysis, we consider multiple ARMA specifications because, for example, a model that fits in-sample data well will not necessarily provide good out-of-sample forecasts.

Figure 4: Actual vs estimated positivity rate

Comparison of actual (black) and in-sample prediction of the Lasso-HAR (red) and of ARIMA(2,1,2) (green) for the daily positivity rate of Italy.

lag for the LHAR, i.e. (1,7,19,25). It is worth noting that in seven out of nine cases the loss functions considered are minimized by HAR-type models. In particular, the LHAR minimizes MAE , MAE_{sd} , MAE_{prop} , MSE , MSE_{sd} whereas the HAR(1,7,14,21,28) minimizes MSE_{prop} and $QLIKE$. Finally, the ARIMA(2,1,2) is the specification that minimizes MAE_{log} and MSE_{log} (even though for MAE_{log} the LHAR returns a very similar result).

In Figure 4, we plot the actual (black line) and the estimated daily PPR given by the LHAR model (red line) and ARIMA(2,1,2) model (green line). It can easily be seen that while the LHAR better follows the dynamics of the percentage positive rate in both low and high infection periods, the ARIMA(2,1,2), being smoother, is not able to fully capture all variations and peaks in the actual PPR, especially in periods characterized by a high viral spread rate. These considerations remain essentially valid also for the other HAR and ARIMA specifications, which have not been shown for ease of interpretation.

4.2. Out-of-sample results

To investigate the predictive ability of the models, we conduct an h -step-ahead rolling window study at the forecasting horizons $h = 1$, $h = 3$ and $h = 7$. The forecasts are obtained by re-estimating the model parameters at each step with a rolling window of 200 observations (2/3 of the sample). To compare the forecasting performances, we consider the set of loss function specified in Section 4, while to assess the significance of differences of the competing models we refer to the MCS relying on the semi-quadratic statistic and the confidence levels of 75% and 90%¹⁰.

¹⁰We report results only for the semi-quadratic statistic because the results corresponding to the range statistic remain substantially stable.

Table 3: Out-of-sample model comparison for Italy: forecast horizon $h = 1$

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	0.7483	0.1335	0.1116	0.1136	0.9387	0.0268	0.0213	0.0226	2.9366
ARIMA(2,1,0)	0.7177	0.1284	0.1078	0.1107	0.8721	0.0250	0.0201	0.0221	2.9361
ARIMA(3,1,0)	0.7192	0.1290	0.1089	0.1126	0.8639	0.0249	0.0202	0.0230	2.9363
ARIMA(0,1,1)	0.7336	0.1285	0.1044	0.1074	0.9261	0.0253	0.0181	0.0202	2.9352
ARIMA(0,1,2)	0.7203	0.1276	0.1056	0.1089	0.8815	0.0248	0.0189	0.0214	2.9356
ARIMA(0,1,3)	0.7263	0.1299	0.1090	0.1127	0.8815	0.0252	0.0200	0.0228	2.9362
ARIMA(1,1,1)	0.7412	0.1319	0.1099	0.1136	0.9170	0.0260	0.0205	0.0234	2.9365
ARIMA(1,1,2)	0.7136	0.1277	0.1074	0.1106	0.8664	0.0247	0.0196	0.0221	2.9359
ARIMA(1,1,3)	0.6970	0.1272	0.1094	0.1142	0.7597	0.0233	0.0202	0.0236	2.9364
ARIMA(2,1,1)	0.7020	0.1255	0.1052	0.1084	0.8392	0.0241	0.0192	0.0215	2.9357
ARIMA(2,1,2)	0.6791	0.1279	0.1181	0.1428	0.7415	0.0270	0.0420	0.1564	2.9566
ARIMA(2,1,3)	0.5881	0.1063	0.0900	0.0940	0.6132	0.0186	0.0156	0.0187	2.9340
ARIMA(3,1,1)	0.7059	0.1261	0.1058	0.1091	0.8703	0.0246	0.0192	0.0218	2.9358
ARIMA(3,1,2)	0.6278	0.1127	0.0951	0.0983	0.6987	0.0205	0.0168	0.0191	2.9345
ARIMA(3,1,3)	0.6696	0.1234	0.1115	0.1352	0.7339	0.0257	0.0393	0.1471	2.9548
HAR(1,7,14)	0.6311	0.1144	0.0981	0.1063	0.6375	0.0205	0.0191	0.0257	2.9363
HAR(1,7,21)	0.6392	0.1144	0.0965	0.1039	0.6699	0.0204	0.0178	0.0233	2.9355
HAR(1,7,28)	0.6480	0.1157	0.0978	0.1049	0.6863	0.0202	0.0174	0.0225	2.9353
HAR(1,7,14,21)	0.6026	0.1091	0.0934	0.1002	0.5786	0.0187	0.0175	0.0230	2.9353
HAR(1,7,14,28)	0.5932	0.1081	0.0940	0.1002	0.5619	0.0180	0.0171	0.0220	2.9351
HAR(1,7,14,21,28)	0.5849	0.1063	0.0921	0.0977	0.5326	0.0170	0.0162	0.0207	2.9346
LHAR	0.5681	0.0955	0.0755	0.0763	0.6324	0.0148	0.0095	0.0101	2.9307
FHAR	0.6006	0.1002	0.0776	0.0789	0.7267	0.0171	0.0105	0.0112	2.9312

The table reports the average values of the different loss functions for the models under analysis. Bold numbers indicate the best performing model by each criterion at the forecast horizon $h = 1$. The numbers shaded in gray and light-gray denote that the corresponding models are included in the 75% and 90% MCS, respectively. We use a rolling window of 200 observations to estimate the coefficients of the models at each step.

We first consider the case of one-day-ahead PPR forecasts ($h = 1$), showing the results in Table 3. Overall, HAR-type models provide significantly better performance than ARIMA models. The lowest loss values are always returned by the LHAR with the single exception of the MSE minimized by the HAR(1,7,14,21,28). The superiority of the HAR models is also confirmed by the MCS as the only ARIMAs falling in the 75% MCS are the ARIMA(2,1,3) for MAE and MAE_{sd} and the ARIMA(3,1,2) for MAE . Some other ARIMA specifications of order $(p, q) \geq 2$ are included in the less restrictive 90% MCS in a few isolated cases. On the other hand, the LHAR and FHAR are the only models always entering the set of superior models for all the forecast criteria considered.

For the remaining forecast horizons, ARIMA models perform poorly in general. Table 4 and Table 5 report the forecast performance of all considered models for $h = 3$ and $h = 7$ periods ahead, respectively. It can easily be seen that the superiority of the HAR models remains unchanged over longer forecast horizons. Among the seven HAR-type models, the LHAR shows

Table 4: Out-of-sample model comparison for Italy: forecast horizon $h = 3$

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	0.8582	0.1506	0.1231	0.1280	1.2388	0.0331	0.0228	0.0253	2.9234
ARIMA(2,1,0)	0.8329	0.1464	0.1197	0.1247	1.1604	0.0315	0.0221	0.0249	2.9231
ARIMA(3,1,0)	0.8301	0.1450	0.1177	0.1226	1.1544	0.0308	0.0210	0.0238	2.9226
ARIMA(0,1,1)	0.8420	0.1450	0.1148	0.1204	1.2400	0.0325	0.0207	0.0238	2.9225
ARIMA(0,1,2)	0.8223	0.1437	0.1164	0.1215	1.1453	0.0309	0.0211	0.0240	2.9226
ARIMA(0,1,3)	0.8361	0.1465	0.1190	0.1243	1.1734	0.0318	0.0219	0.0250	2.9231
ARIMA(1,1,1)	0.8375	0.1455	0.1169	0.1220	1.2006	0.0320	0.0214	0.0242	2.9228
ARIMA(1,1,2)	0.8257	0.1459	0.1198	0.1240	1.1535	0.0315	0.0220	0.0247	2.9230
ARIMA(1,1,3)	0.7875	0.1404	0.1159	0.1214	0.9991	0.0285	0.0206	0.0238	2.9224
ARIMA(2,1,1)	0.8085	0.1416	0.1151	0.1196	1.0867	0.0294	0.0205	0.0231	2.9223
ARIMA(2,1,2)	0.8370	0.1471	0.1202	0.1249	1.1488	0.0312	0.0218	0.0245	2.9229
ARIMA(2,1,3)	0.8294	0.1419	0.1119	0.1158	1.1731	0.0301	0.0196	0.0220	2.9218
ARIMA(3,1,1)	0.8314	0.1451	0.1178	0.1218	1.2191	0.0319	0.0216	0.0242	2.9228
ARIMA(3,1,2)	0.7975	0.1387	0.1131	0.1173	1.0853	0.0282	0.0203	0.0232	2.9222
ARIMA(3,1,3)	0.8148	0.1404	0.1123	0.1161	1.1414	0.0299	0.0206	0.0234	2.9224
HAR(1,7,14)	0.6210	0.1168	0.1052	0.1152	0.6091	0.0203	0.0202	0.0265	2.9227
HAR(1,7,21)	0.6635	0.1204	0.1042	0.1129	0.7229	0.0213	0.0185	0.0233	2.9216
HAR(1,7,28)	0.7076	0.1259	0.1064	0.1144	0.7999	0.0221	0.0175	0.0213	2.9210
HAR(1,7,14,21)	0.5718	0.1068	0.0957	0.1030	0.4940	0.0167	0.0167	0.0211	2.9207
HAR(1,7,14,28)	0.5628	0.1037	0.0916	0.0975	0.4868	0.0155	0.0146	0.0177	2.9194
HAR(1,7,14,21,28)	0.5510	0.1002	0.0873	0.0924	0.4641	0.0142	0.0131	0.0157	2.9186
LHAR	0.5274	0.0851	0.0639	0.0645	0.6050	0.0130	0.0071	0.0075	2.9153
FHAR	0.5660	0.0941	0.0737	0.0757	0.6215	0.0139	0.0086	0.0095	2.9161

The table reports the average values of the different loss functions for the models under analysis. Bold numbers indicate the best performing model by each criterion at the forecast horizon $h = 3$. The numbers shaded in gray and light-gray denote that the corresponding models are included in the 75% and 90% MCS, respectively. We use a rolling window of 200 observations to estimate the coefficients of the models at each step.

a dominant position for predicting the PPR at $h = 3$ since it minimizes the loss functions in eight out of nine cases and it is the only model that always enters 75% MCS (Table 4). At the same time, the FHAR provides better forecast accuracy at a weekly ($h = 7$) horizon, achieving the lowest losses for all criteria used and being the only model permanently included in 75% MCS (Table 5).

Summarizing, the out-of-sample results clearly show that HAR-type specifications outperform ARIMA models in predicting the positivity rate at the considered forecast horizons $h = 1$, $h = 3$ and $h = 7$. Also, it should be noted that, overall, both Lasso-HAR and Flexible-HAR, taking into account uncertainty in model specification, outperform the standard HAR based on a fixed lag index at each forecast horizon and by each criterion. Therefore, the above results suggest that allowing a more general specification of HAR is successful probably because the lasso-based models include only active predictors, letting the lag structure approximate the long memory observed in the data.

Table 5: Out-of-sample model comparison for Italy: forecast horizon $h = 7$

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	0.9970	0.1770	0.1433	0.1590	1.8202	0.0526	0.0350	0.0473	2.9011
ARIMA(2,1,0)	0.9801	0.1738	0.1404	0.1561	1.7838	0.0517	0.0342	0.0464	2.9006
ARIMA(3,1,0)	0.9523	0.1687	0.1360	0.1512	1.7092	0.0497	0.0328	0.0447	2.8999
ARIMA(0,1,1)	0.9943	0.1749	0.1392	0.1554	1.8495	0.0532	0.0344	0.0470	2.9008
ARIMA(0,1,2)	0.9706	0.1712	0.1371	0.1529	1.7725	0.0514	0.0337	0.0462	2.9004
ARIMA(0,1,3)	0.9649	0.1707	0.1371	0.1530	1.7539	0.0513	0.0339	0.0465	2.9005
ARIMA(1,1,1)	0.9783	0.1725	0.1379	0.1538	1.8020	0.0519	0.0339	0.0462	2.9005
ARIMA(1,1,2)	0.9284	0.1626	0.1310	0.1434	1.6679	0.0461	0.0311	0.0414	2.8988
ARIMA(1,1,3)	1.0900	0.1884	0.1471	0.1634	2.0626	0.0575	0.0361	0.0490	2.9017
ARIMA(2,1,1)	0.9096	0.1602	0.1290	0.1420	1.5421	0.0443	0.0297	0.0397	2.8981
ARIMA(2,1,2)	0.9830	0.1742	0.1405	0.1560	1.7774	0.0521	0.0350	0.0481	2.9011
ARIMA(2,1,3)	0.8779	0.1567	0.1295	0.1425	1.4197	0.0417	0.0302	0.0403	2.8983
ARIMA(3,1,1)	0.9871	0.1679	0.1325	0.1446	1.8515	0.0479	0.0313	0.0419	2.8989
ARIMA(3,1,2)	0.8999	0.1564	0.1279	0.1377	1.4222	0.0376	0.0276	0.0359	2.8967
ARIMA(3,1,3)	0.9989	0.1684	0.1328	0.1401	1.7923	0.0441	0.0294	0.0352	2.8972
HAR(1,7,14)	0.7186	0.1402	0.1335	0.1486	0.8704	0.0290	0.0295	0.0386	2.8979
HAR(1,7,21)	0.7906	0.1479	0.1341	0.1474	1.0414	0.0308	0.0271	0.0343	2.8964
HAR(1,7,28)	0.8304	0.1511	0.1324	0.1454	1.1411	0.0324	0.0268	0.0342	2.8963
HAR(1,7,14,21)	0.6713	0.1290	0.1206	0.1316	0.6812	0.0229	0.0234	0.0296	2.8944
HAR(1,7,14,28)	0.6569	0.1234	0.1129	0.1228	0.6609	0.0217	0.0221	0.0284	2.8937
HAR(1,7,14,21,28)	0.6600	0.1221	0.1105	0.1195	0.6579	0.0208	0.0210	0.0266	2.8931
LHAR	0.5815	0.0925	0.0677	0.0697	0.7312	0.0162	0.0096	0.0111	2.8868
FHAR	0.5156	0.0848	0.0647	0.0671	0.5763	0.0138	0.0090	0.0106	2.8865

The table reports the average values of the different loss functions for the models under analysis. Bold numbers indicate the best performing model by each criterion at the forecast horizon $h = 7$. The numbers shaded in gray and light-gray denote that the corresponding models are included in the 75% and 90% MCS, respectively. We use a rolling window of 200 observations to estimate the coefficients of the models at each step.

4.3. Robustness check: out-of-sample results for the United States

We investigate the robustness of the proposed approach by considering the evolution of PPR for the United States (US). The 2019 coronavirus pandemic has led to massive social upheaval around the world and in the US. As mentioned above, the first outbreak of the virus occurred in the city of Wuhan in China's Hubei province in December 2019, but the virus then spread to Asia, Europe and North America between January and March 2020. By the end of March, there were more than 700,000 confirmed cases of Covid-19 worldwide and more than 34,000 people had died from causes related to the virus, with the US reaching more confirmed cases than any other country, surpassing China and Italy with more than 86,000 positive tests (Johns Hopkins University, 2020).

The data to conduct our analysis on the PPR for the US were downloaded from the *Our World in Data* website¹¹ by Roser et al. (2020). Since data on testing are not available for the early

¹¹<https://ourworldindata.org/coronavirus>

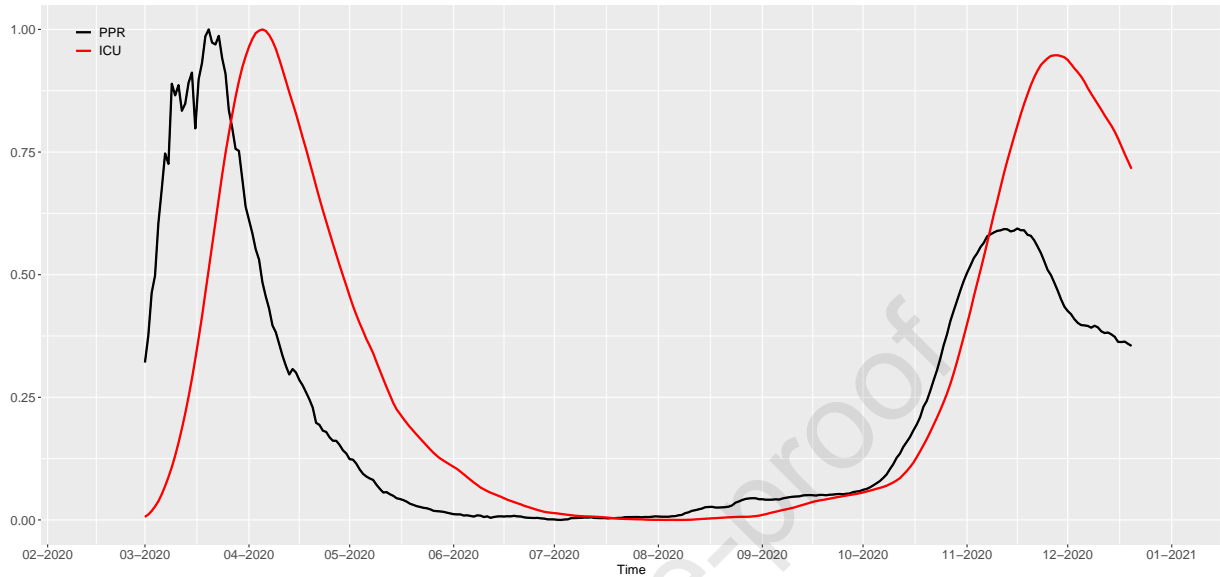
phase of the pandemic the sample period goes from 01 March to 20 December 2020. As with the Italian PPR, forecasts are obtained by recursively estimating model parameters every day over a 200-day rolling window. Accordingly, the out-of-sample period runs from 16 September to 20 December 2020. However, in order to improve the overall presentation of the paper, the tables with the forecasting results have been reported in the Appendix, while we will discuss only the main findings here. Considering the forecasting horizons $h = 1$, $h = 3$ and $h = 7$, the results for the US confirm what was found for Italy. In particular, it turns out that LHAR and FHAR are the only models that consistently enter the MCS regardless of the chosen forecast horizon, always minimizing each of the nine loss functions considered. For $h = 1$ and $h = 3$ no other models are included in the MCS, while for $h = 7$ some HAR specifications with fixed lag index enter the set of superior models but only for MAE-type loss functions. Overall, for short forecast horizons, ARMA-type and HAR-type models with fixed-based lag structure tend to have similar performance, but for longer forecast horizons HAR specifications prevail. On the other hand, the lasso-based HAR specifications outperform competing models in each forecasting scenario, capturing both short- and long-run PPR dynamics.

5. Discussion

As discussed above, Covid-19 is an infection characterized by a high percentage of asymptomatic cases. Several studies have shown that more than 40% of cases may not reveal symptoms. This means that no country knows the true total number of people affected by Covid-19, but all we know is the infection status of who has been tested. As a result, testing is a crucial element in understanding the spread of the ongoing pandemic (Hasell et al., 2020). Better than simply counting the total number of tests and in conjunction with data on confirmed cases, the positivity rate represents a key metric for understanding the pandemic, as it measures both the severity of the epidemic and the limitations of testing. According to WHO, before a country can loosen restrictions or begin reopening, the positivity rate for a comprehensive testing program should be 5% or less for at least 14 days.

High rates of positivity occur when, for example, the only people being tested are patients with more severe clinical symptoms who are suspected of having Covid-19 and have required hospitalization. Consequently, a high PPR means that countries should probably aim for a larger and more comprehensive testing program, suggesting that it is not a reasonable time to relax restrictions designed to reduce coronavirus transmission. At the same time, because a high positive rate suggests high rates of infection in the community caused by rapid transmission of the virus, this indicates that it may be useful to impose restrictions to slow the spread of the disease. Achieving a low test positivity rate may be the result of a large enough testing volume such that asymptomatic and mild cases as well as exposed contacts are monitored, even if they are asymptomatic. On the other hand, low positivity rates can be the result of enacting different types of public health interventions such as encouraging smart working, closing schools, banning mass gatherings, restrictions on eating in public places along with permissive to total lockdown home stay measures.

Such a situation requires the parallel development of at least two dimensions: on the one hand, policymakers should plan ahead for needs in terms of medical facilities and equipment, while, on the other hand, analytical tools and models allowing the generation of reliable forecasts and future scenarios should be developed. For example, analyzing the data in Figure 5, which shows the time series of the normalized 7-day moving average of PPR and patients in Intensive

Figure 5: Positivity rate vs hospitalized ICU patients

Comparison of 7-day moving average of positivity rate in black and hospitalized patients in intensive care units (ICU) in red for Italy. For ease of comparison, variables were normalized to have a scale between 0 and 1.

Care Unit (ICU) for Italy, it turns out that the peak of patients in ICU occurs between 12 and 16 days after the peak of PPR. Similar scenarios also arise at the regional level. (Fenga and Gaspari, 2021). In this perspective, the approach proposed in this paper could help decision makers to plan public health policies in advance because the PPR could have a predictive capacity with respect to hospitalizations, changing the level of intensity of these interventions over the course of the epidemic. In addition, the ability of HAR to reproduce the short- and medium/long-term trend of the positivity rate could avoid both the immediate economic costs of lockdown and the societal costs of social distancing measures. For example, a long-term upward prediction of the PPR, could lead to a testing strategy on a larger population scale to monitor and reduce viral transmission. This will inevitably also have repercussions on the social networks of the various actors in the supply chain.

6. Conclusion

Understanding the dynamics of the current epidemic is essential for the development of non-pharmaceutical interventions and thus for reducing the health, economic and social impacts caused by the pandemic. This paper focuses on the positivity rate as it represents a crucial metric for understanding the Covid-19 outbreak. The positivity rate offers a measure of how adequately countries are testing and provides insight on the current level of coronavirus transmission in the community. Since ARMA models have been found to be poorly suited to model the long-term behavior of time series, in this paper, we propose to use the HAR approach to model the slowly-moving long-run level and the highly persistent autocorrelation structure that characterize the Covid-19 positivity rate.

The empirical study of the Italian positivity rate, along with robustness checks on the US data, shows that HAR models generally outperform ARIMA specifications under various criteria and forecast horizons. The forecasting superiority of the HAR emerges from the MCS where the standard HAR and its lasso-based alternatives significantly outperform ARIMA models under the forecast horizons $h = 1$, $h = 3$ and $h = 7$. In particular, the gains widen as the forecast horizon increases. Also, the out-of-sample results point out that the more general HAR lasso-based lag structure is preferable compared with the HAR fixed-based lag structure. These results are confirmed by the in-sample analysis, as allowing for model specification uncertainty under the HAR framework leads to improvements in model fitting, minimizing the loss functions considered.

Thus, this approach is particularly useful as it allows for accurately forecasting short-, medium-, and long-term trends of the positivity rate. **In this regard, monitoring the trend in positivity rate and ICU admissions suggests that the PPR might have predictive ability with respect to hospitalizations because peaks in positivity rate precede peaks in hospitalizations, which occur on average within an interval of 12 to 16 days later.** Generating accurate forecasts at different horizons of the PPR is relevant for reducing uncertainty around interventions, leading to maximize resource and investment allocations. Therefore, understanding the trend in Covid-19 positivity rates would allow governments to modify their social and health policies in advance. Also, since the model components are chosen in a completely data-driven fashion, this significantly reduces the uncertainty and arbitrariness associated with model specification. In this respect, lasso-based HAR-type models simplify the decision-making process by leading it towards a common direction driven by the dynamics of the data.

However, it is worth noting that policy decisions should not only be determined by the positivity rate, but also by other health, economic, demographic, environmental, and climate variables. **This is because the positivity rate provides some useful information about testing capacity and the spread of the virus in the community, but it can also depend on factors such as how it is calculated, the testing accessibility and the timeliness of the laboratory in providing results. Of course, no single metric gives us a complete picture of the prevalence of Covid-19 in the community. Therefore, each week, we need to monitor PPR trends along with other metrics such as recovered and active cases, percent change in new cases, hospitalizations and deaths.**

In the United States, there are no federal standards for reporting Covid-19 test data. It makes impossible to offer a single view of testing data at the national level and consequently test data are reported differently. In addition, there are several possible ways to calculate test positivity. For example, on the Johns Hopkins University & Medicine webpage *Differences in Positivity Rates* four possible ways to calculate positivity rates are outlined.

Regarding Italy, the Italian Civil Protection provides comprehensive data at the regional level on some variables of interest such as swabs analyzed, cumulative confirmed cases, home isolation cases, hospitalized cases, ICU cases, and deaths. However, regionalization of the health care system and data fragmentation pose challenges in the management of the Covid-19 outbreak in Italy. This has led to the enactment of several regional policies, especially in terms of testing strategies. Consequently, in order to trace the real extension of Covid-19 infection, official data must be interpreted with caution considering several aspects. For example, because there are some inconsistencies and delays in the transmission of this data, on some days negative values of positive cases, tests, and deaths are reported. In addition, short-term fluctuations could affect the reliability of daily data. These fluctuations may be the result of laboratory delays (laboratory

saturation) or calendar effects (typically, the number of tests tends to decrease on weekends). Thus, these inconsistencies should be considered to account for data variability. In this direction, in addition to official national bulletins, it might be useful to cross-reference information from different data sources.

Currently, a complete picture of the drivers of Covid-19 spread that clarifies the causes of the variability in infections across provinces and regions within countries is still lacking. Although human-to-human transmission is recognized as the primary vehicle for virus transmission, several studies have argued that virus circulation may also be associated with geographical, environmental, and socio-economic factors. Accordingly, using information on these factors could allow the definition and refinement of epidemiological modelling and thus the design of appropriate policy responses to manage this threat to population health and, more generally, to socio-economic systems. In this perspective, the proposed approach could be improved by including health, environmental, and socio-economic variables in the HAR models.

Along these lines, as a direction for future research, it could be useful to investigate, and potentially include in HAR models, the relationships between positivity rate and epidemiological variables (effective reproductive number R_t); demographic parameters (social interactions, age, and sex); environmental and climatic factors (humidity, wind speed, and temperature); pollution indicators (air quality); and socio-economic activities (economic and social interactions within and between countries). At the same time, it might be appropriate to consider data inconsistencies. Considering these variables when generating forecasts could provide decision makers with better guidance in establishing additional control measures, loosening restrictions, enhancing benefits, and preventing the failure of measures already taken.

Acknowledgment

I would like to thank the Editor, the Associate Editor and two anonymous Referees for their highly constructive comments which make this paper more valuable.

Disclosure statement

The author has no conflicts of interest to disclose.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

- Albani, V. V., R. M. Velho, and J. P. Zubelli (2021). Estimating, monitoring, and forecasting covid-19 epidemics: a spatiotemporal approach applied to nyc data. *Scientific Reports* 11(1), 1–15.
- Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885–905.
- Audrino, F., C. Huang, and O. Okhrin (2019). Flexible har model for realized volatility. *Studies in Nonlinear Dynamics & Econometrics* 23(3), 20170080.
- Audrino, F. and S. D. Knaus (2016). Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews* 35(8-10), 1485–1521.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73(1), 5–59.
- Baker, S. R., N. Bloom, S. J. Davis, K. Kost, M. Sammon, and T. Viratyosin (2020). The unprecedented stock market reaction to covid-19. *The review of asset pricing studies* 10(4), 742–758.

- Bashir, M. F., B. Ma, B. Komal, M. A. Bashir, D. Tan, M. Bashir, et al. (2020). Correlation between climate indicators and covid-19 pandemic in new york, usa. *Science of The Total Environment* 728, 138835.
- Benvenuto, D., M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi (2020). Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, 105340.
- Berardi, C., M. Antonini, M. G. Genie, G. Cotugno, A. Lanteri, A. Melia, and F. Paolucci (2020). The covid-19 pandemic in italy: Policy and technology impact on health and non-health outcomes. *Health policy and technology* 9(4), 454–487.
- Bertuzzo, E., L. Mari, D. Pasetto, S. Miccoli, R. Casagrandi, M. Gatto, and A. Rinaldo (2020). The geography of covid-19 spread in italy and implications for the relaxation of confinement measures. *Nature communications* 11(1), 1–11.
- Bloise, F. and M. Tancioni (2021). Predicting the spread of covid-19 in italy using machine learning: Do socio-economic factors matter? *Structural Change and Economic Dynamics* 56, 310–329.
- Bontempi, E. and M. Coccia (2021). International trade as critical parameter of covid-19 spread that outclasses demographic, economic, environmental, and pollution factors. *Environmental Research*, 111514.
- Bontempi, E., M. Coccia, S. Vergalli, and A. Zanoletti (2021). Can commercial trade represent the main indicator of the covid-19 diffusion due to human-to-human interactions? a comparative analysis between italy, france, and spain. *Environmental Research*, 111529.
- Cao, L.-t., H.-h. Liu, J. Li, X.-d. Yin, Y. Duan, and J. Wang (2020). Relationship of meteorological factors and human brucellosis in hebei province, china. *Science of The Total Environment* 703, 135491.
- Ceylan, Z. (2020). Estimation of covid-19 prevalence in italy, spain, and france. *Science of The Total Environment*, 138817.
- Chowdhury, E. K., I. I. Khan, and B. K. Dhar (2021). Catastrophic impact of covid-19 on the global stock markets and economic activities. *Business and Society Review*.
- Chowdhury, M. T., A. Sarkar, S. K. Paul, and M. A. Moktadir (2020). A case study on strategies to deal with the impacts of covid-19 pandemic in the food and beverage industry. *Operations Management Research*, 1–13.
- Chowdhury, M. T., A. Sarkar, P. K. Saha, and R. H. Anik (2020). Enhancing supply resilience in the covid-19 pandemic: a case study on beauty and personal care retailers. *Modern Supply Chain Research and Applications*.
- Coccia, M. (2021). The relation between length of lockdown, numbers of infected people and deaths of covid-19, and economic growth of countries: Lessons learned to cope with future pandemics similar to covid-19 and to constrain the deterioration of economic system. *Science of The Total Environment* 775, 145801.
- Copat, C., A. Cristaldi, M. Fiore, A. Grasso, P. Zuccarello, S. Santo Signorelli, G. O. Conti, and M. Ferrante (2020). The role of air pollution (pm and no₂) in covid-19 spread and lethality: a systematic review. *Environmental research*, 110129.
- Copiello, S. and C. Grillenzoni (2020). The spread of 2019-ncov in china was primarily driven by population density. comment on “association between short-term exposure to air pollution and covid-19 infection: Evidence from china” by zhu et al. *Science of The Total Environment* 744, 141028.
- Cori, A. (2020). *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*. R package version 2.2-3.
- Cori, A., N. M. Ferguson, C. Fraser, and S. Cauchemez (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* 178(9), 1505–1512.
- Corsi, F. (2009, 02). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7(2), 174–196.
- De Silva, U. S. K., A. Paul, K. W. Hasan, S. K. Paul, S. M. Ali, and R. K. Chakraborty (2021). Examining risks and strategies for the spice processing supply chain in the context of an emerging economy. *International Journal of Emerging Markets*.
- Deaton, B. J. and B. J. Deaton (2020). Food security and canada’s agricultural system challenged by covid-19. *Canadian Journal of Agricultural Economics/Revue canadienne d’agroeconomie* 68(2), 143–149.
- Delamater, P. L., E. J. Street, T. F. Leslie, Y. T. Yang, and K. H. Jacobsen (2019). Complexity of the basic reproduction number (r_0). *Emerging infectious diseases* 25(1), 1.
- Dente, S. and S. Hashimoto (2020). Covid-19: A pandemic with positive and negative outcomes on resource and waste flows and stocks. *Resources, Conservation and Recycling* 161, 104979.
- Diao, Y., S. Koder, D. Anzai, J. Gomez-Tames, E. A. Rashed, and A. Hirata (2021). Influence of population density, temperature, and absolute humidity on spread and decay durations of covid-19: A comparative study of scenarios in china, england, germany, and japan. *One Health* 12, 100203.
- Du, Z., X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers (2020). Serial interval of covid-19 among publicly reported confirmed cases. *Emerging infectious diseases* 26(6), 1341.
- Earnest, A., M. I. Chen, D. Ng, and L. Y. Sin (2005). Using autoregressive integrated moving average (arima) models to predict and monitor the number of beds occupied during a sars outbreak in a tertiary hospital in singapore. *BMC Health Services Research* 5(1), 36.
- Ehlert, A. (2021). The socio-economic determinants of covid-19: A spatial analysis of german county level data. *Socio-*

- Economic Planning Sciences* 78, 101083.
- Eslami, H. and M. Jalili (2020). The role of environmental factors to transmission of sars-cov-2 (covid-19). *Amb Express* 10(1), 1–8.
- Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96(456), 1348–1360.
- Fasina, F. O., M. A. Salami, M. M. Fasina, O. A. Otekunrin, A. L. Hoogesteijn, and J. B. Hittner (2021). Test positivity – evaluation of a new metric to assess epidemic dispersal mediated by non-symptomatic cases. *Methods*.
- Fattorini, D. and F. Regoli (2020). Role of the chronic air pollution levels in the covid-19 outbreak risk in italy. *Environmental Pollution* 264, 114732.
- Fenga, L. and M. Gaspari (2021). Predictive capacity of covid-19 test positivity rate. *Sensors* 21(7).
- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* 33(1), 1.
- Gaudart, J., O. Touré, N. Dessay, A. Iassane Dicko, S. Ranque, L. Forest, J. Demongeot, and O. K. Doumbo (2009). Modelling malaria incidence with environmental dependency in a locality of sudanese savannah area, mali. *Malaria journal* 8(1), 61.
- Giordano, G., F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri (2020). Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature medicine* 26(6), 855–860.
- Gray, R. S. (2020). Agriculture, transportation, and the covid-19 crisis. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie* 68(2), 239–243.
- Gunessee, S. and N. Subramanian (2020). Ambiguity and its coping mechanisms in supply chains lessons from the covid-19 pandemic and natural disasters. *International Journal of Operations & Production Management*.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- Hasell, J., E. Mathieu, D. Beltekian, B. Macdonald, C. Giattino, E. Ortiz-Ospina, M. Roser, and H. Ritchie (2020). A cross-country database of covid-19 testing. *Scientific data* 7(1), 1–7.
- He, Z. and H. Tao (2018). Epidemiology and arima model of positive-rate of influenza viruses among children in wuhan, china: A nine-year retrospective study. *International Journal of Infectious Diseases* 74, 61–70.
- Ibn-Mohammed, T., K. Mustapha, J. Godsell, Z. Adamu, K. Babatunde, D. Akintade, A. Acquaye, H. Fujii, M. Ndiaye, F. Yamoah, and S. Koh (2021). A critical analysis of the impacts of covid-19 on the global economy and ecosystems and opportunities for circular economy strategies. *Resources, Conservation and Recycling* 164, 105169.
- Ivanov, D. (2020). Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the covid-19 pandemic. *Annals of Operations Research*, 1–21.
- Ivanov, D. and A. Das (2020). Coronavirus (covid-19/sars-cov-2) and supply chain resilience: A research note. *International Journal of Integrated Supply Management* 13(1), 90–102.
- Ivanov, D. and A. Dolgui (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. a position paper motivated by covid-19 outbreak. *International Journal of Production Research* 58(10), 2904–2915.
- Jayaweera, M., H. Perera, B. Gunawardana, and J. Manatunge (2020). Transmission of covid-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy. *Environmental research* 188, 109819.
- Johns Hopkins University (2020). Coronavirus resource center. <https://coronavirus.jhu.edu/>. Accessed: December 2020.
- Klose, J. and P. Tillmann (2021). Covid-19 and financial markets: A panel analysis for european countries. *Jahrbücher für Nationalökonomie und Statistik* 241(3), 297–347.
- Koonin, L. M. (2020). Novel coronavirus disease (covid-19) outbreak: Now is the time to refresh pandemic plans. *Journal of business continuity & emergency planning* 13(4), 298–312.
- Liu, Y., S. Funk, S. Flasche, et al. (2020). The contribution of pre-symptomatic infection to the transmission dynamics of covid-2019. *Wellcome Open Research* 5(58), 58.
- Lozano-Diez, J. A., J. A. Marmolejo-Saucedo, and R. Rodriguez-Aguilar (2020, 1). Designing a resilient supply chain: An approach to reduce drug shortages in epidemic outbreaks. *EAI Endorsed Transactions on Pervasive Health and Technology* 6(21).
- Majumdar, A., M. Shaw, and S. K. Sinha (2020). Covid-19 debunks the myth of socially sustainable supply chain: A case of the clothing industry in south asian countries. *Sustainable Production and Consumption* 24, 150–155.
- Monllor, P., Z. Su, L. Gabrielli, and P. Taltavull de La Paz (2020). Covid-19 infection process in italy and spain: Are data talking? evidence from arma and vector autoregression models. *Frontiers in Public Health* 8, 784.
- Müller, U. A., M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen, and J. R. Ward (1993). Fractals and intrinsic time—a challenge to econometricians. *Unpublished manuscript, Olsen & Associates, Zürich*, 130.
- Murgante, B., G. Borruso, G. Balletto, P. Castiglia, and M. Dettori (2020). Why italy first? health, geographical and

- planning aspects of the covid-19 outbreak. *Sustainability* 12(12), 5064.
- Nishiura, H., N. M. Linton, and A. R. Akhmetzhanov (2020). Serial interval of novel coronavirus (covid-19) infections. *International journal of infectious diseases*.
- Onder, G., G. Rezza, and S. Brusaferro (2020). Case-fatality rate and characteristics of patients dying in relation to covid-19 in italy. *Jama* 323(18), 1775–1776.
- Oran, D. P. and E. J. Topol (2020). Prevalence of asymptomatic sars-cov-2 infection: a narrative review. *Annals of internal medicine* 173(5), 362–367.
- Panarello, D. and G. Tassinari (2021). One year of covid-19 in italy: are containment policies enough to shape the pandemic pattern? *Socio-Economic Planning Sciences*, 101120.
- Patton, A., D. N. Politis, and H. White (2009). Correction to “automatic block-length selection for the dependent bootstrap” by d. politis and h. white. *Econometric Reviews* 28(4), 372–375.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160(1), 246–256.
- Paul, S. K. and P. Chowdhury (2020a). A production recovery plan in manufacturing supply chains for a high-demand item during covid-19. *International Journal of Physical Distribution & Logistics Management*.
- Paul, S. K. and P. Chowdhury (2020b). Strategies for managing the impacts of disruptions during covid-19: an example of toilet paper. *Global Journal of Flexible Systems Management* 21(3), 283–293.
- Porta, M. (2014). *A dictionary of epidemiology*. Oxford university press.
- Queiroz, M. M., D. Ivanov, A. Dolgui, and S. F. Wamba (2020). Impacts of epidemic outbreaks on supply chains: mapping a research agenda amid the covid-19 pandemic through a structured literature review. *Annals of operations research*, 1–38.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Remko, V. H. (2020). Research opportunities for a more resilient post-covid-19 supply chain—closing the gap between research findings and industry practice. *International Journal of Operations & Production Management* 40(4), 341–355.
- Rendana, M. et al. (2021). Air pollutant levels during the large-scale social restriction period and its association with case fatality rate of covid-19. *Aerosol and Air Quality Research* 21, 200630–200630.
- Roser, M., H. Ritchie, E. Ortiz-Ospina, and J. Hasell (2020). Coronavirus pandemic (covid-19). *Our world in data*.
- Sahai, A. K., N. Rath, V. Sood, and M. P. Singh (2020). Arima modelling & forecasting of covid-19 in top five affected countries. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14(5), 1419–1427.
- Sarkis, J. (2020). Supply chain sustainability: learning from the covid-19 pandemic. *International Journal of Operations & Production Management*.
- Sarkis, J., M. J. Cohen, P. Dewick, and P. Schröder (2020). A brave new world: Lessons from the covid-19 pandemic for transitioning to sustainable supply and production. *Resources, Conservation, and Recycling* 159, 104894.
- Sharma, M., S. Luthra, S. Joshi, and A. Kumar (2020). Developing a framework for enhancing survivability of sustainable supply chains during and post-covid-19 pandemic. *International Journal of Logistics Research and Applications* 0(0), 1–21.
- Singh, R. K., M. Rani, A. S. Bhagavathula, R. Sah, A. J. Rodriguez-Morales, H. Kalita, C. Nanda, S. Sharma, Y. D. Sharma, A. A. Rabaan, et al. (2020). Prediction of the covid-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (arima) model. *JMIR public health and surveillance* 6(2), e19115.
- Srivastava, A. (2020). Covid-19 and air pollution and meteorology—an intricate relationship: A review. *Chemosphere*, 128297.
- Sun, J., X. Chen, Z. Zhang, S. Lai, B. Zhao, H. Liu, S. Wang, W. Huan, R. Zhao, M. T. A. Ng, et al. (2020). Forecasting the long-term trend of covid-19 epidemic using a dynamic model. *Scientific reports* 10(1), 1–10.
- Taqi, H. M., H. N. Ahmed, S. Paul, M. Garshasbi, S. M. Ali, G. Kabir, S. K. Paul, et al. (2020). Strategies to manage the impacts of the covid-19 pandemic in the supply chain: implications for improving economic and social sustainability. *Sustainability* 12(22), 9483.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–288.
- van Barneveld, K., M. Quinlan, P. Kriesler, A. Junor, F. Baum, A. Chowdhury, P. N. Junankar, S. Clibborn, F. Flanagan, C. F. Wright, et al. (2020). The covid-19 pandemic: Lessons on building more equal and sustainable societies. *The Economic and Labour Relations Review* 31(2), 133–157.
- Verikios, G. (2020). The dynamic effects of infectious disease outbreaks: The case of pandemic influenza and human coronavirus. *Socio-Economic Planning Sciences* 71, 100898.
- Wallinga, J. and P. Teunis (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology* 160(6), 509–516.

- World Health Organization (2020). Public health criteria to adjust public health and social measures in the context of covid-19: annex to considerations in adjusting public health and social measures in the context of covid-19, 12 may 2020. Technical documents.
- Zhang, D., M. Hu, and Q. Ji (2020). Financial markets under the global pandemic of covid-19. *Finance Research Letters* 36, 101528.
- Zheng, Y.-L., L.-P. Zhang, X.-L. Zhang, K. Wang, and Y.-J. Zheng (2015). Forecast model analysis for the morbidity of tuberculosis in xinjiang, china. *PloS one* 10(3), e0116832.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association* 101(476), 1418–1429.

Appendix A. Out-of-sample results for the United States

Table A.1: Out-of-sample model comparison for United States: forecast horizon $h = 1$

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	1.0553	0.1906	0.1432	0.1453	1.9594	0.0597	0.0333	0.0362	2.9967
ARIMA(2,1,0)	1.0415	0.1870	0.1394	0.1428	1.9562	0.0595	0.0330	0.0372	2.9968
ARIMA(3,1,0)	0.9598	0.1735	0.1303	0.1341	1.7499	0.0539	0.0305	0.0344	2.9956
ARIMA(0,1,1)	1.0284	0.1858	0.1395	0.1423	1.8471	0.0575	0.0330	0.0357	2.9966
ARIMA(0,1,2)	0.9287	0.1670	0.1245	0.1283	1.7151	0.0517	0.0282	0.0317	2.9944
ARIMA(0,1,3)	0.9273	0.1681	0.1265	0.1307	1.6695	0.0514	0.0290	0.0330	2.9948
ARIMA(1,1,1)	0.9415	0.1707	0.1284	0.1322	1.7033	0.0525	0.0296	0.0334	2.9951
ARIMA(1,1,2)	0.9324	0.1682	0.1258	0.1298	1.6920	0.0514	0.0284	0.0322	2.9945
ARIMA(1,1,3)	0.9212	0.1674	0.1264	0.1314	1.6168	0.0494	0.0277	0.0327	2.9944
ARIMA(2,1,1)	0.9217	0.1671	0.1258	0.1302	1.6718	0.0515	0.0291	0.0332	2.9949
ARIMA(2,1,2)	0.8477	0.1554	0.1184	0.1221	1.4082	0.0447	0.0260	0.0294	2.9933
ARIMA(2,1,3)	0.7434	0.1380	0.1066	0.1100	1.2843	0.0410	0.0243	0.0275	2.9924
ARIMA(3,1,1)	0.8437	0.1542	0.1171	0.1216	1.4783	0.0463	0.0267	0.0306	2.9937
ARIMA(3,1,2)	0.7659	0.1422	0.1097	0.1135	1.2933	0.0417	0.0248	0.0282	2.9927
ARIMA(3,1,3)	0.7232	0.1352	0.1052	0.1082	1.1834	0.0382	0.0230	0.0256	2.9917
HAR(1,7,14)	0.8247	0.1494	0.1126	0.1152	1.4253	0.0435	0.0244	0.0277	2.9925
HAR(1,7,21)	0.8187	0.1485	0.1121	0.1145	1.4267	0.0435	0.0244	0.0276	2.9925
HAR(1,7,28)	0.8181	0.1486	0.1122	0.1149	1.4355	0.0437	0.0246	0.0280	2.9926
HAR(1,7,14,21)	0.8153	0.1480	0.1116	0.1142	1.4322	0.0438	0.0247	0.0280	2.9926
HAR(1,7,14,28)	0.8134	0.1476	0.1114	0.1143	1.4247	0.0433	0.0243	0.0278	2.9924
HAR(1,7,14,21,28)	0.8069	0.1461	0.1100	0.1133	1.4092	0.0428	0.0239	0.0278	2.9923
LHAR	0.5638	0.1019	0.0767	0.0789	0.7843	0.0225	0.0119	0.0140	2.9861
FHAR	0.5723	0.1035	0.0779	0.0801	0.8314	0.0240	0.0126	0.0148	2.9865

The table reports the average values of the different loss functions for the models under analysis. Bold numbers indicate the best performing model by each criterion at the forecast horizon $h = 1$. The numbers shaded in gray and light-gray denote that the corresponding models are included in the 75% and 90% MCS, respectively. We use a rolling window of 200 observations to estimate the coefficients of the models at each step.

Table A.2: Out-of-sample model comparison for United States: forecast horizon $h = 3$

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	1.3046	0.2393	0.1820	0.1867	2.4443	0.0790	0.0465	0.0496	3.0001
ARIMA(2,1,0)	1.2713	0.2336	0.1780	0.1837	2.2935	0.0742	0.0438	0.0485	2.9991
ARIMA(3,1,0)	1.1647	0.2164	0.1667	0.1727	1.9047	0.0644	0.0396	0.0440	2.9970
ARIMA(0,1,1)	1.2249	0.2244	0.1704	0.1748	2.1210	0.0683	0.0401	0.0431	2.9970
ARIMA(0,1,2)	0.9900	0.1838	0.1413	0.1475	1.5059	0.0502	0.0303	0.0343	2.9923
ARIMA(0,1,3)	0.9888	0.1850	0.1433	0.1500	1.5240	0.0518	0.0319	0.0362	2.9932
ARIMA(1,1,1)	1.0341	0.1923	0.1481	0.1540	1.6032	0.0538	0.0326	0.0363	2.9934
ARIMA(1,1,2)	0.9946	0.1851	0.1426	0.1491	1.5201	0.0510	0.0310	0.0351	2.9927
ARIMA(1,1,3)	0.9562	0.1782	0.1378	0.1446	1.4728	0.0499	0.0310	0.0360	2.9928
ARIMA(2,1,1)	0.9989	0.1882	0.1469	0.1537	1.5532	0.0539	0.0339	0.0384	2.9942
ARIMA(2,1,2)	1.0193	0.1922	0.1501	0.1566	1.5953	0.0557	0.0349	0.0399	2.9947
ARIMA(2,1,3)	0.8696	0.1661	0.1315	0.1358	1.2422	0.0449	0.0294	0.0321	2.9917
ARIMA(3,1,1)	0.9590	0.1810	0.1416	0.1483	1.4256	0.0501	0.0320	0.0363	2.9932
ARIMA(3,1,2)	0.9208	0.1755	0.1385	0.1439	1.2725	0.0467	0.0308	0.0344	2.9925
ARIMA(3,1,3)	0.8532	0.1635	0.1299	0.1335	1.1287	0.0419	0.0281	0.0305	2.9910
HAR(1,7,14)	0.7215	0.1322	0.1003	0.1028	0.8750	0.0280	0.0163	0.0179	2.9850
HAR(1,7,21)	0.7298	0.1340	0.1021	0.1042	0.8932	0.0287	0.0167	0.0182	2.9852
HAR(1,7,28)	0.7363	0.1351	0.1028	0.1051	0.9122	0.0290	0.0168	0.0184	2.9853
HAR(1,7,14,21)	0.7186	0.1317	0.1000	0.1025	0.8899	0.0284	0.0165	0.0181	2.9851
HAR(1,7,14,28)	0.7189	0.1315	0.0998	0.1023	0.8879	0.0280	0.0161	0.0178	2.9849
HAR(1,7,14,21,28)	0.7167	0.1313	0.0997	0.1024	0.8696	0.0275	0.0158	0.0176	2.9848
LHAR	0.4436	0.0813	0.0618	0.0634	0.4057	0.0126	0.0071	0.0078	2.9803
FHAR	0.4404	0.0808	0.0614	0.0629	0.4032	0.0126	0.0071	0.0077	2.9803

The table reports the average values of the different loss functions for the models under analysis. Bold numbers indicate the best performing model by each criterion at the forecast horizon $h = 3$. The numbers shaded in gray and light-gray denote that the corresponding models are included in the 75% and 90% MCS, respectively. We use a rolling window of 200 observations to estimate the coefficients of the models at each step.

Table A.3: Out-of-sample model comparison for United States: forecast horizon $h = 7$

	MAE	MAE_{sd}	MAE_{log}	MAE_{prop}	MSE	MSE_{sd}	MSE_{log}	MSE_{prop}	$QLIKE$
ARIMA(1,1,0)	0.9863	0.1846	0.1431	0.1519	1.4816	0.0500	0.0305	0.0359	2.9797
ARIMA(2,1,0)	0.8793	0.1626	0.1243	0.1315	1.2286	0.0397	0.0229	0.0268	2.9758
ARIMA(3,1,0)	0.7504	0.1413	0.1098	0.1160	0.8809	0.0301	0.0183	0.0212	2.9733
ARIMA(0,1,1)	0.8573	0.1591	0.1220	0.1292	1.1286	0.0374	0.0221	0.0259	2.9753
ARIMA(0,1,2)	0.7070	0.1315	0.1008	0.1068	0.7684	0.0256	0.0150	0.0175	2.9716
ARIMA(0,1,3)	0.7597	0.1411	0.1080	0.1146	0.8444	0.0281	0.0164	0.0191	2.9724
ARIMA(1,1,1)	0.7598	0.1418	0.1090	0.1163	0.8761	0.0296	0.0176	0.0209	2.9730
ARIMA(1,1,2)	0.7374	0.1369	0.1047	0.1111	0.8195	0.0272	0.0158	0.0184	2.9720
ARIMA(1,1,3)	0.8600	0.1650	0.1310	0.1403	1.0296	0.0382	0.0255	0.0305	2.9772
ARIMA(2,1,1)	0.7417	0.1388	0.1072	0.1136	0.8044	0.0273	0.0164	0.0189	2.9723
ARIMA(2,1,2)	0.7811	0.1444	0.1101	0.1161	0.9999	0.0324	0.0185	0.0215	2.9734
ARIMA(2,1,3)	0.8545	0.1576	0.1198	0.1275	1.1900	0.0387	0.0221	0.0266	2.9754
ARIMA(3,1,1)	0.7080	0.1330	0.1030	0.1089	0.7435	0.0254	0.0153	0.0176	2.9717
ARIMA(3,1,2)	0.8488	0.1551	0.1169	0.1238	1.1641	0.0363	0.0200	0.0237	2.9743
ARIMA(3,1,3)	0.8424	0.1558	0.1189	0.1270	1.1784	0.0386	0.0223	0.0270	2.9756
HAR(1,7,14)	0.4955	0.0920	0.0701	0.0728	0.4496	0.0144	0.0079	0.0090	2.9679
HAR(1,7,21)	0.5184	0.0957	0.0727	0.0751	0.4744	0.0149	0.0082	0.0091	2.9680
HAR(1,7,28)	0.5283	0.0970	0.0731	0.0756	0.4825	0.0150	0.0081	0.0091	2.9680
HAR(1,7,14,21)	0.5155	0.0951	0.0721	0.0749	0.4703	0.0147	0.0080	0.0090	2.9679
HAR(1,7,14,28)	0.5121	0.0939	0.0707	0.0732	0.4618	0.0143	0.0077	0.0086	2.9677
HAR(1,7,14,21,28)	0.5049	0.0928	0.0701	0.0725	0.4498	0.0139	0.0075	0.0084	2.9676
LHAR	0.4638	0.0848	0.0639	0.0655	0.3661	0.0114	0.0062	0.0068	2.9669
FHAR	0.4732	0.0866	0.0651	0.0669	0.3862	0.0120	0.0066	0.0071	2.9671

The table reports the average values of the different loss functions for the models under analysis. Bold numbers indicate the best performing model by each criterion at the forecast horizon $h = 7$. The numbers shaded in gray and light-gray denote that the corresponding models are included in the 75% and 90% MCS, respectively. We use a rolling window of 200 observations to estimate the coefficients of the models at each step.

Highlights

- The Covid-19 pandemic is severely affecting public health and global economies.
- We propose the HAR approach for fitting and forecasting the Covid-19 positivity rate.
- HAR models outperform ARIMA in predicting positivity rates at different horizons.
- Accurately predicting positivity rates helps in planning socio-economic strategies.
- Our findings are supported by an empirical application to Italian and US data.

Antonio Naimoli received a Ph.D. in Financial Econometrics from the University of Salerno, Department of Economics and Statistics. He is currently Research Associate in Economic Statistics at the same Department. His research interests are mainly related to the modelling and forecasting of financial volatility and trading volumes. He has published papers on related topics in international journals such as *International Journal of Forecasting*, *Quantitative Finance*, *Economic Modelling* and *Journal of Risk and Financial Management*. He has also refereed articles for *Bulletin of Economic Research*, *Journal of Risk and Financial Management* and *Applied Sciences*.

Antonio Naimoli
Dept. of Economics and Statistics (DISES)
University of Salerno
Via G. Paolo II 132, 84084

The article contains original research and is not considered for publication elsewhere.

This paper is submitted with full responsibility, following the due ethical procedures. There is no duplicate publication, fraud or plagiarism.

I would also like to point out that there are no conflicts of interest with other persons or organizations that could inappropriately influence the content of the article.